Best Practices for the HP EVA Array using VMware vCenter Site Recovery Manager

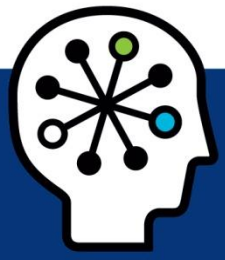


Table of contents

Introduction 2
HP StorageWorks Continuous Access EVA 3
Data replication 3
DR groups and copy sets 3
Write history log (log disk) 3
Continuous Access failover 3

Solution Scope 4
Software and hardware configuration 6
Configuration details 6
VMware vCenter Server setup 6
Basic VMware ESX Server installation 6
SAN switch zoning 6
Configure EVAs at source and destination sites 7
EVA disk group capacity planning 8
Configure DR groups in HP Command View 9
VMware vCenter Site Recovery Manager and Continuous Access EVA 9
Setting up the Storage Replication Adapter and Command View EVA 9
Site Recovery Manager Objects relationship 15
Configuring data replication 16

Optimizing performance 20
ESX SAN load balancing and HP StorageWorks EVA 20
Continuous Access replication tunnels 24
Monitoring EVA activity 25
Sizing the SAN ISL 26

Testing 26
Performance testing 28
DR group full copy performance 29

Failover considerations 31

Failback considerations 33

Best practices and caveats 34
ESX server and Site Recovery Manager administration 34
Storage administration 35
Implementation caveats 35

Conclusion 36

For more information 37

Introduction

VMware deployments are increasingly being used for mission-critical applications. In response to this trend, VMware delivered VMware vCenter Site Recovery Manager, which works in conjunction with storage array replication software to construct, manage, maintain, test, and execute a fully automated disaster recovery plan for VMware virtual machines and their data.

This paper describes a set of detailed best practices to plan, setup, deploy, and operate a remote copy infrastructure using Site Recovery Manager in conjunction with HP StorageWorks Continuous Access EVA Software and the HP EVA Storage Replication Adapter. A building block approach is used to design the complete solution, providing a step-by-step set of best practices to deploy the requisite technical components from VMware and HP, showing the relationships and interdependencies.

The following fundamental deployment issues that this paper addresses are as follows:

- Operating VMware virtual machines with Continuous Access EVA
- Deploying representative workloads in remote copied VMs comprised of various industry applications, such as Microsoft Exchange, Microsoft SQL server, and File and Print services.
- Creating a representative workflow (recovery plan) that fails-over various sets of VMs to the recovery site
- Operating the workflow in a test scenario
- Operating the workflow for a full failover

Developing best practices and providing steps to show how to properly fail-back from the recovery site to the protected site

The methods and best practices described in following sections are designed to facilitate both the productive planning and the timely deployment of a fully automated disaster recovery plan, using Site Recovery Manager in conjunction with a combination of HP c-Class Blades and EVA arrays using Continuous Access. The following guidelines and best practices ensure these benefits:

- Proper configuration and deployment of the entire environment, including the VMware environment, the HP Blade Servers, the EVA Disk Array, and the EVA Site Recovery Adapter.
- Predictable performance characteristics
- Ease of operation, including tips for how to properly interface HP technology with VMware technology.

Use these best practices to accelerate time to deployment, while reducing risks and minimizing total costs.

HP StorageWorks Continuous Access EVA

Continuous Access EVA is a feature of the Enterprise Virtual Array (EVA) that allows data replication between two or more EVAs. This section describes some basic Continuous Access EVA terminology, concepts, and features. The following topics are discussed.

- Data replication
- Copy sets
- Data replication (DR) groups
- Log disk
- Failover

Data replication

Data replication with Continuous Access EVA is host-independent and supports various interconnection technologies, such as Fibre Channel over IP (FCIP) and Fibre Channel. Additionally, the EVA also supports bidirectional replication. When a storage system contains both source virtual disks (Vdisks) and destination Vdisks, it is bidirectional. A given storage system can have a bidirectional data replication relationship with only one other storage system, and an individual Vdisk can have a unidirectional-replicating relationship with only one other Vdisk. Continuous Access EVA enables data replication between all models of the EVA family. Continuous Access EVA can replicate data synchronously and asynchronously between source and destination arrays. Data replication between sites is most widely used when creating a true disaster-tolerant data center as described in this document. It can also migrate data between two EVA arrays or provide an alternative method for performing backups.

DR groups and copy sets

A copy set is a generic term for a replicated Vdisk. A data replication (DR) group is comprised of copy sets (replicated Vdisks). Each DR group acts as a consistency group—all of its member Vdisks replicate to the same destination EVA, failover together, preserve write order across members, and share a write history log. Therefore, a DR group is the primary level of Continuous Access management.

Write history log (log disk)

The DR group has storage allocated on demand for its write history log (WHL). The WHL collects hosts write commands and data if access to the destination storage system is unavailable. When the connection is re-established, the content of the WHL is written to the destination Vdisk for quick resynchronization. Log writes occur in the same order that the host writes occurred, this process is called merging.

Continuous Access failover

The recovery process whereby one or more DR group switches over to its backup is called a failover. The process can be planned or unplanned. A planned failover allows an orderly shutdown of the attached hosts and the EVA controllers before the redundant system takes over. An unplanned failover occurs when a failure or outage occurs that may not allow an orderly transition of roles.

Solution Scope

Building an effective disaster tolerant solution can often be a very complex and time consuming task. Furthermore most disaster tolerant solutions, implemented at customer sites, are often untested and may fail to protect customers when failure occurs. Depending on the data center solution or application, the recovery point objective and recovery time objective may differ from customer to customer. Any disaster tolerant solution must be able to accommodate both planned and unplanned downtime.

Planned downtime

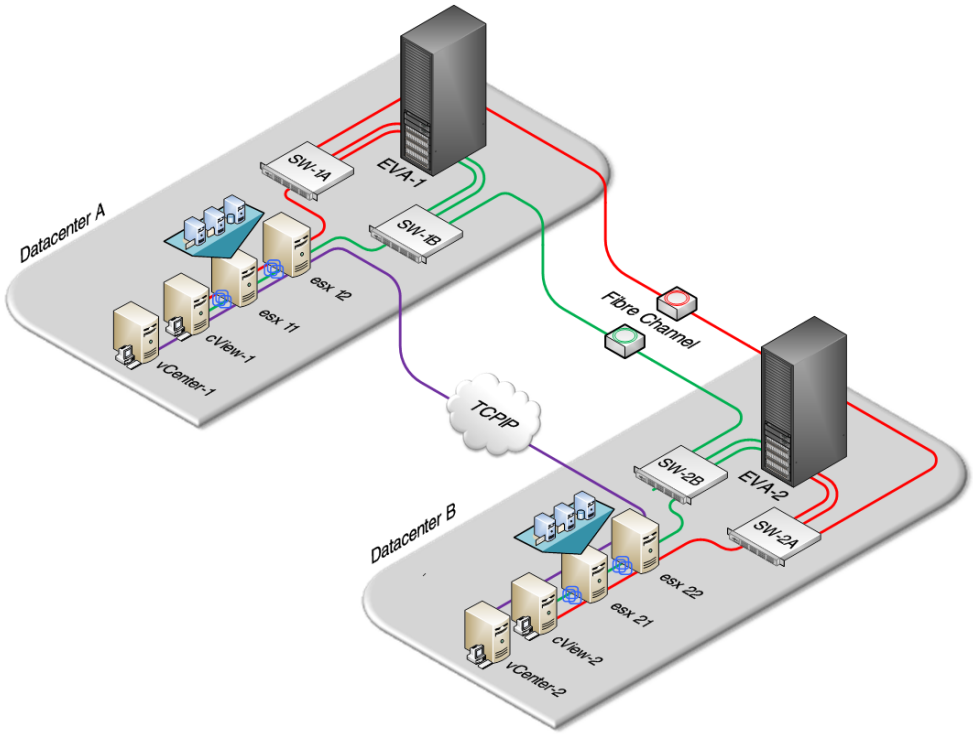
Planned downtime is a result of equipment of software maintenance that is disruptive to system operation and usually cannot be avoided with a currently installed system design. In general, planned downtime is the result of some logical, management-initiated event.

Unplanned downtime

Unplanned downtime results from a natural disaster, environmental anomaly or sudden hardware or software failure and may impact an entire site. HP StorageWorks Continuous Access EVA and VMware vCenter Site Recovery Manager offer an ideal solution, allowing you to move your data and secure your mission critical applications to a safe offsite data center, away from harm.

This case study demonstrates a disaster-tolerant solution between two data centers in a campus or metropolitan area. In this scenario, the data centers are merged through a shared dual redundant storage area network (SAN) (see Figure 1). EVA 1 and EVA 2 are connected through the shared SAN and data replicated synchronously with Continuous Access.

Figure 1. Configuration diagram



In this configuration, the VMware ESX servers can be clustered so that failures of a single ESX host only require that VMs be restarted on another host in the same datacenter. For more information about VMware High Availability (HA) see the VMware HA Concepts, Implementation and Best Practices white paper.

Failures of an EVA array require a failover of Continuous Access DR groups and a restart of the VMs in the other data center. These operations are handled by VMware vCenter Site Recovery Manager. This solution supports active/active replication by which each data center backs up the other. Both sides are protected and recovery data centers.

Our disaster recovery objectives are defined as follows:

- Recovery Point Objective: Low
- Recovery Time Objective: Medium

Low=0 to seconds
Medium=minute to half hour
High=half hour to several hours

Software and hardware configuration

The following software and hardware components are used in this case study.

Software

- VMware ESX server 3.5 U4
- VMware vCenter 2.5 U4
- VMware vCenter Site Recovery Manager 1.0 U1
- HP Site Recovery Adapter for EVA 1.01
- HP StorageWorks Command View EVA 8.0.1 or 9.1

Hardware

- HP StorageWorks Enterprise Virtual Array, configured with:
    - HP Fibre Channel (FC) Drives
    - HP StorageWorks Continuous Access
    - HP StorageWorks Business Copy
- HP ProLiant BL480c and BL460c Servers
- Brocade 4/24 Storage Area Network (SAN) Switch for HP c-Class BladeSystem
- Cisco Catalyst Blade Switch 3020 for HP c-Class BladeSystem

Configuration details

VMware vCenter Server setup

- Set up vCenter Server at each site
- Create a single data center in each instance of vCenter Server
- Add the local hosts to this data center

Basic VMware ESX Server installation

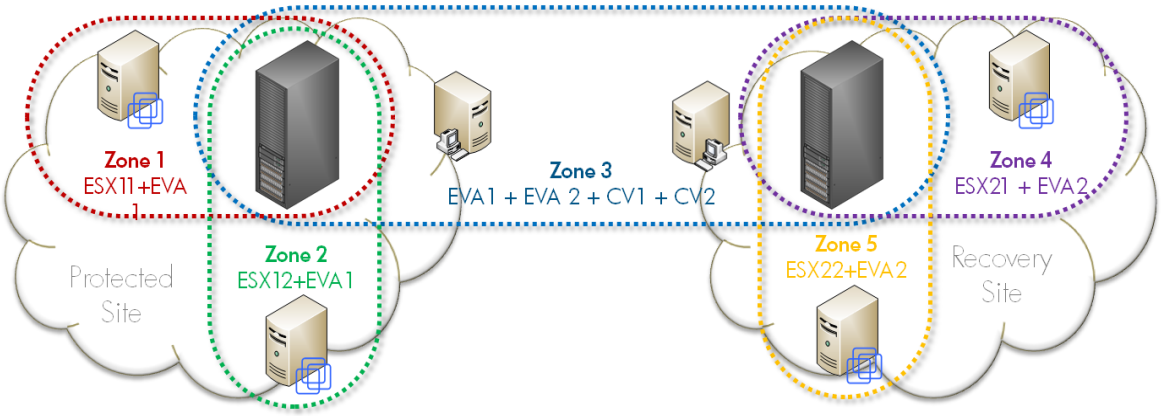
Follow your best practices for installing and configuring VMware Infrastructure 3. The ESX host can either boot from local disks or from the SAN.

SAN switch zoning

Configure zoning in a way that all ESX servers for a given data center see only the local EVA. When zoning is properly configured, ESX servers (esx11, esx12) in Data Center A have access to the local storage (EVA 1) in Data Center A, and ESX servers (esx21, esx22) in Data Center B have access to the local storage (EVA 2) in Data Center B.

Just as you would do in other environments, follow the SAN best practices. HP implements a best practice to create separate single initiator/single target zones. (For example, limit the creation of two zones per ESX Cluster in each fabric, as shown in Figure 2). This configuration provides a shared SAN environment in which the FC switch from Data Center A and Data Center B are merged through inter-switch links (ISL) into one SAN. For more information about zoning and SAN design considerations, see the *HP SAN design reference guide*.

Figure 2. Zoning example



SAN switch settings for Continuous Access

We used the following fabric Brocade Switch settings, which are based on rules detailed in the SAN design guide and the HP StorageWorks HP StorageWorks Continuous Access EVA Implementation Guide.

- `aptpolicy` 1 to set port-based routing policy
- `iodset` to guarantee in-order delivery
- `dlsreset` to disable dynamic path selection

Configure EVAs at source and destination sites

Planning EVA disk groups

Planning the necessary disk groups to meet your I/O requirements should be done when configuring the array. See the HP StorageWorks Enterprise Virtual Array configuration best practices white paper for recommendations on configuring the array properly.

Determining the number of disks in a disk group

Performance and capacity are the most important factors when designing a storage solution, because they drive the number of physical disks required for the solution. Performance is a function of how quickly a transaction can be completed by the storage system. For transactional applications like MS Exchange 2007 or SQL server, which typically generate small block random IOPs, a goal of 20 ms is used for database reads and writes, while 10 ms is used for log writes. See the HP Best practices for deploying an Exchange 2007 environment using VMware ESX 3.5 white paper for recommendations on configuring Microsoft Exchange 2007 and the EVA array properly.

Determining how many spindles are required for random access performance begins with understanding the following workload characteristics.

- The total IOPs required by the applications
- The Read/Write ratios (for example, 1:1 for Exchange 2007 and about 2:1 for a generic OLTP application on SQL Server 2008)

These characteristics must then be matched with the raw performance of the physical disks that are being used (for example, 170 IOPS per 15 K RPM disk or 125 IOPS per 10 K RPM disk with small block random workloads) and the desired RAID protection.

To determine the number of spindles required for performance, the following simple formula can be used:

$$\text{Number of drives needed} = \frac{(\text{Total IOPs} * \text{RAID penalty} * \text{write \%}) + (\text{Total IOPs} * \text{read \%})}{\text{Raw performance of the disk drive}}$$

The RAID penalty represents the I/O overhead associated with a RAID level; that is, the number of I/Os to the physical disks incurred by a write operation coming in from a host. Table 1 lists the penalties for every supported RAID level on the EVA.

Table 1. RAID-level penalties

Raid Level	I/O Penalty
RAID 0	1
RAID 1	2
RAID 5	4
RAID 6	6

Note: When data is replicated remotely, application performance is not necessarily improved by increasing the number of disks in a disk group, because response time for application writes includes the time for replication. In synchronous mode, performance will likely be limited by replication before it is limited by the number of disks.

Note: Sequential access (read or write) is limited by the per-disk performance rather than by the number of disks in the disk group. Consider approx 10 MB/s per physical disk.

EVA disk group capacity planning

After determining the amount of capacity needed for the virtual machines operating system files and data, it is important to understand how much formatted capacity you will get from a physical drive and what RAID overhead will be applied to the group of disks. In the industry, disk drive capacity is expressed and advertised using the decimal representation. This is a source of much confusion since most software including operating systems or the EVA firmware use a binary representation. For example a 300 GB disk once configured in the EVA has a formatted capacity expressed in binary terms of only 279.39 GB or approximately 7% difference.

The EVA uses active sparing for data protection. The spare space for disk failure protection is actually spread across the entire disk group. This provides instant availability to the EVA to transfer data and rebuild a failed drive. Doing so ensures that there is always enough available capacity and no possibility that the disk spare could go bad, leaving you without protection. The sparing capacity equates to two times the largest hard drive in the disk group for single sparing and four times for double sparing.

The following formula is used to determine the number of physical disks needed to deliver the required capacity.

$$\text{Number of drives needed} = \frac{\text{Data Size} + (\text{Data Size} * \text{Raid overhead})}{\text{Formatted Capacity}} + \text{Sparing Capacity}$$

For example, a disk group configuration using 300 GB drives, single sparing, and Vraid 5 for a total estimated VM disk space of 8 TB breaks down as follows:

$$\text{Number of drives needed} = \frac{8000\text{GB}^{\text{DATA}} + (8000\text{GB} * .20^{\text{Vraid 5}})}{279.39\ \text{GB}} + 558.78\text{GB}^{\text{SPARING}} = 37\ \text{Disks}$$

Note: It is a best practice to round up the produce of the performance or capacity sizing formulas to a multiple of 8 disks

Planning for DR group write history log (WHL)

The DR group WHL is a virtual disk that stores a DR group host write data. The log is automatically created when you create the DR group. Once the log is created, it cannot be moved. You must plan for the additional disk capacity required for each DR group WHL. For more information on DR group log size, see the HP StorageWorks Continuous Access EVA Implementation Guide.

Determining the number of disk groups

To determine if the default disk group will meet your remote replication needs, consider the following:

- Separate disk groups can help ensure that data is recoverable if a disk group fails. However, multiple disk groups result in a slightly higher cost of ownership and potentially lower performance.
- In general, distributing the workload across the greatest number of disks in a single disk group provides the best performance. However, separate disk groups can improve performance for sequential workloads such as database logs and rich content.
- Disk groups must provide sufficient free space for snapshots and snapclones (if used), and for Continuous Access DR group WHLs.

Configure DR groups in HP Command View

- On the source EVA, create one or more DR group(s) for all existing ESX Vdisks. Multiple groups should be created to balance the replication load between EVA controllers and Fabric ISLs.
- Leave the default Synchronous write mode and Destination Hosts Access settings at = None
- On the destination EVA (for this test, EVA 2), present all protected Vdisks to the recovery ESX host(s) with the same LUN IDs as are on the source EVA 1.
- Repeat the above actions to configure the EVA array in the remote data center if you are configuring bidirectional replication.

VMware vCenter Site Recovery Manager and Continuous Access EVA

Setting up the Storage Replication Adapter and Command View EVA

VMware vCenter Site Recovery Manager controls and orchestrates failover and test operations on the array thru a software component called the Storage Replication Adapter (SRA).

The HP StorageWorks Command View EVA Software suite is required to manage and configure EVA arrays. It provides centralized management, provisioning, health monitoring, and control of EVA in a SAN. The EVA Storage Replication Adapter (EVA SRA) provides the communication layer between Site Recovery Manager and Command View EVA.

The EVA SRA must be installed on both the local (protected) server and the remote (recovery) Site Recovery Manager servers. For instructions, see the HP StorageWorks EVA Virtualization SRM Adapter administrator guide. It is recommended to deploy the Site Recovery Manager server and vCenter server components into separate machines. However, with small environments (typically less than 40 VMs), the Site Recovery Manager server and vCenter server can reside alongside on the same machine. Similarly, they can share the same SQL or Oracle server instance to house their backend databases. The communication between the Site Recovery Manager Array Manager and the EVA arrays is out-of-band using an SSL network connection

to the Command View server; therefore the Site Recovery Manager server and the HP EVA SRA can be installed either on a physical machine or on a virtual machine.

Note: When using a firewall between the Site Recovery Manager servers and Command View servers, ensure that TCP port 2372 is enabled. This is the default communication port of Command View EVA.

There can be multiple Command View instances in an environment for the purposes of fault tolerance or load balancing. However, there can only be one Command View instance actively managing a given EVA array at any time. Any instance that is not currently actively managing an EVA is considered a passive instance of Command View. A passive instance cannot be used to manage the array without a management role change. This can create a situation where the only Command View server available to Site Recovery Manager is a passive instance. Fortunately, Site Recovery Manager supports multiple Array Manager (SRA instance). EVA SRA 1.01 can automatically activate a passive CV instance. Automatic activation is enabled for failover in test or recovery mode only. However, other Array Manager functions, such as the array identification and LUN inventory, do not trigger an automatic activation. For this reason, the EVA SRA must be configured to use network addresses and credentials for both Command View servers. Addresses, user names, and passwords must be entered without spaces. Each entry must be separated with a semicolon (;).

For example:

- Storage Manager Address: address1;address2
- Username: username1;username2
- Password: password1;password2

To configure Site Recovery Manager to use HP EVA Virtualization Adapter with the Array Manager, open the Add Array Manager window, enter the Command View addresses, usernames, and passwords in the entry fields without spaces, and separate each entry with a semicolon (;). Figure 3 and Figure 4 show screenshots of the Array Manager window.

Figure 3. Protected Site Array Manager configuration

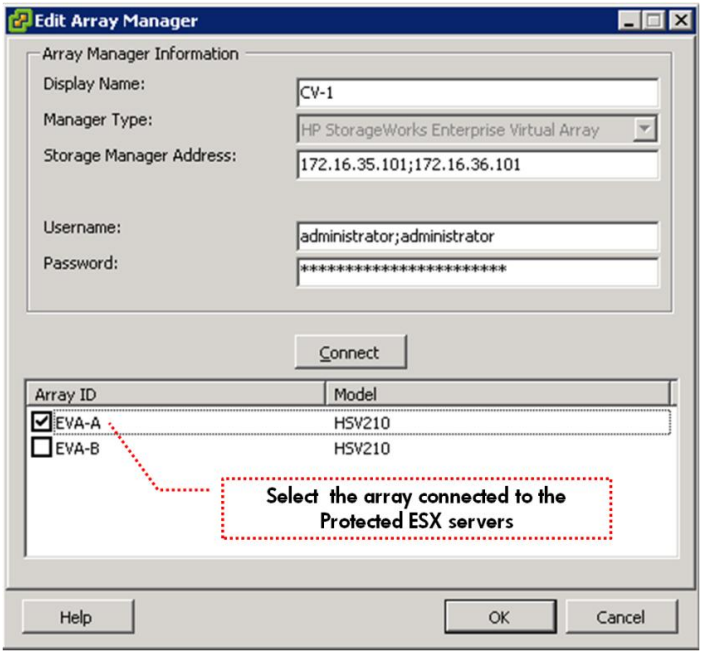
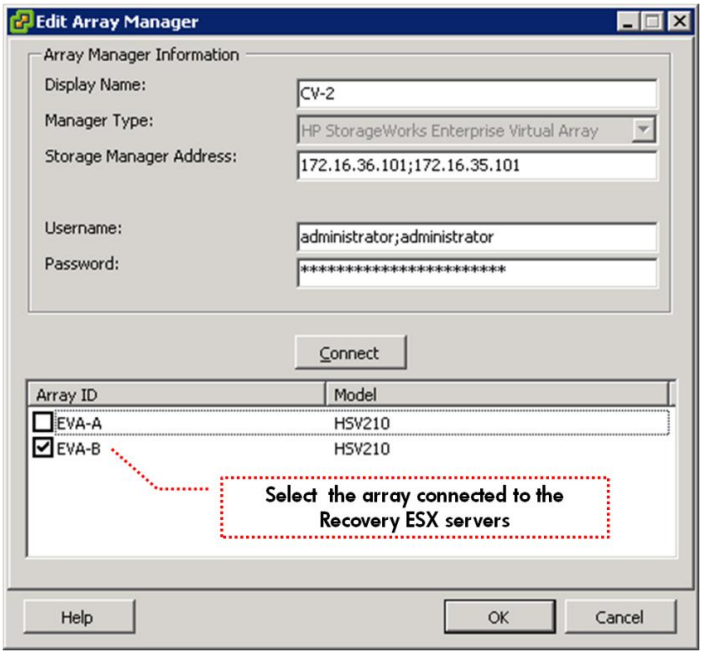
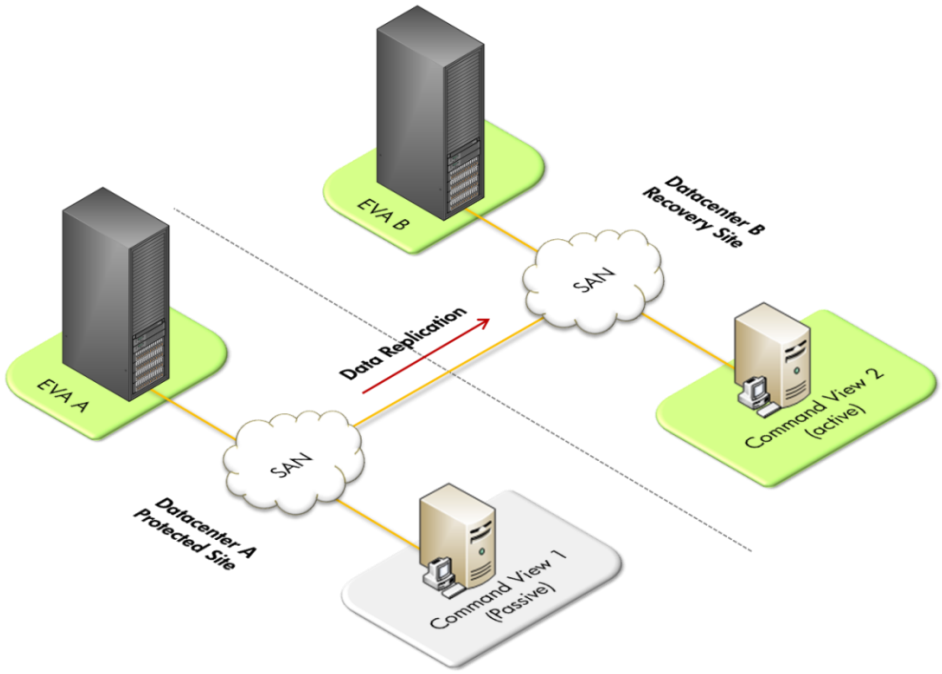


Figure 4. Recovery Side Array Manager configuration



The most common deployment scenario is to have a Command View server at Site A or Site B, managing both the protected and the recovery arrays. Figure 5 shows an Active/Passive Command View deployment. In this example, the instance of Command View (CV-2) placed in the recovery site is actively managing EVA 1 and EVA 2 arrays. In the event of a failover from Site B to Site A (test or recovery mode), the EVA SRA attempts to use CV-2 to execute the storage failover operations. IF CV-2 is not available or not responsive, the EVA SRA automatically activates the passive instance CV-1 to manage the failover.

Figure 5. Active/Passive Command View EVA deployment

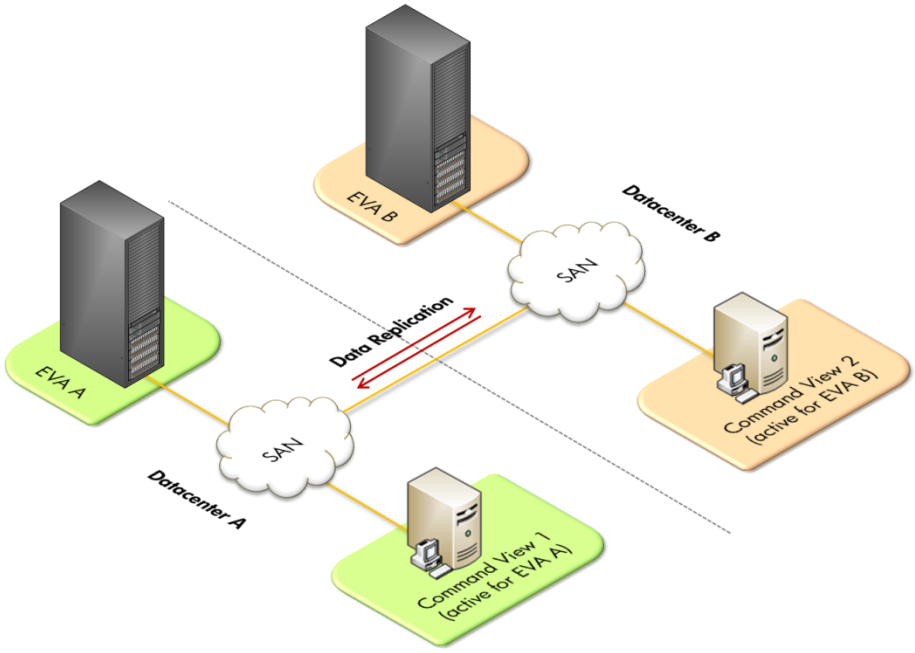


The EVA SRA for Site Recovery Manager also supports active/active Command View deployment. In this scenario, each storage system is managed by a different Command View server. As with active/passive scenarios, an automatic Command View management change occurs when performing a Site Recovery Manager failover in test and recovery mode.

However, this configuration is less practical from a Continuous Access management standpoint. The drawback with an active/active configuration is that some EVA management tasks need both the source and destination arrays to be available to Command View. This is the case, for example, when you create a DR group. As a result, an Active/Active deployment may require the administrator to frequently change the management ownership.

The diagram in Figure 6 represents an active/active configuration in a bidirectional replication solution. The infrastructure at each site is managed by different teams, with each local Command View actively managing the local array. The Site Recovery Manager Array Manager is configured to use CV1 and CV2.

Figure 6. Active/Active Command View placement

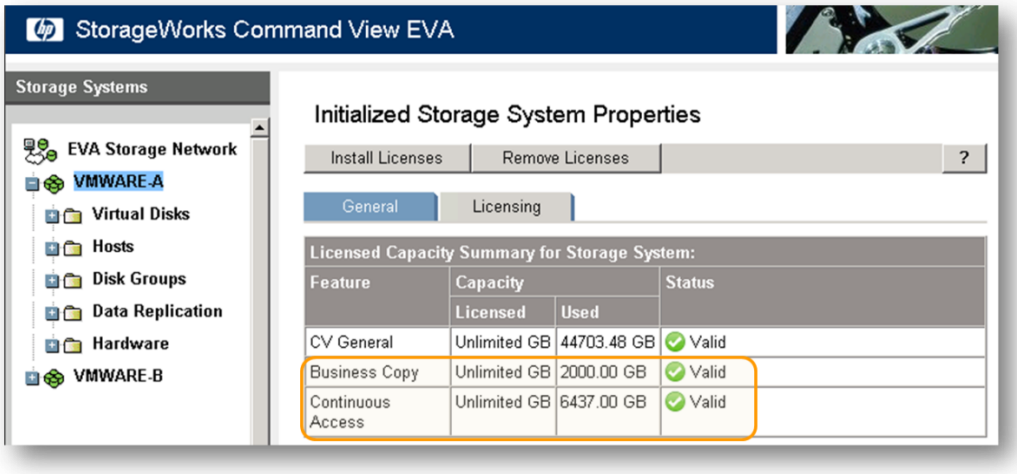


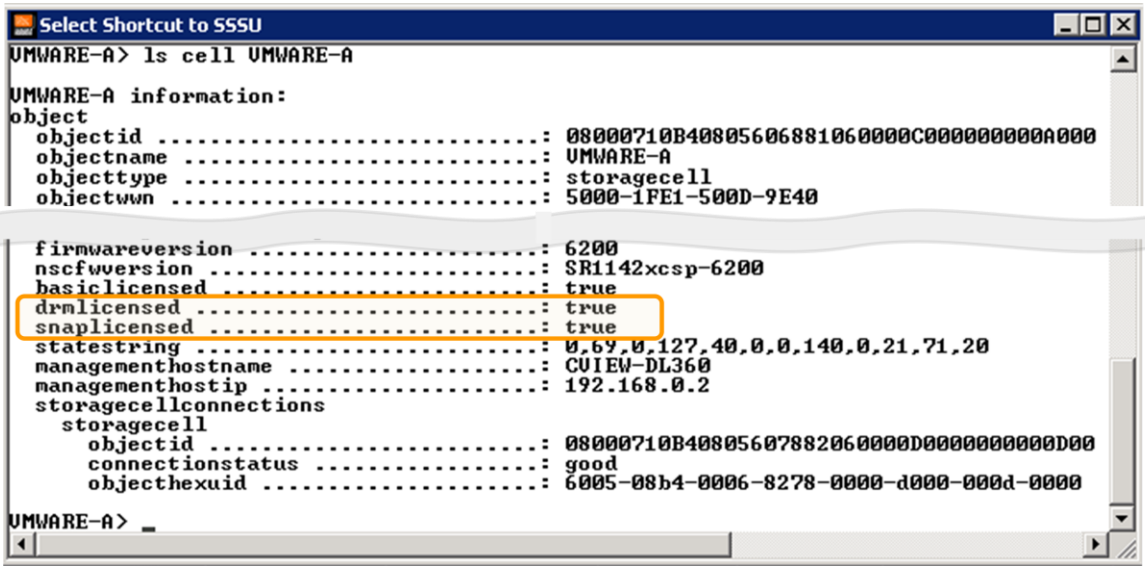
Installing EVA licenses

HP StorageWorks Continuous Access EVA is needed to create replication pairs and perform failover between data centers. HP StorageWorks Business Copy EVA is needed to create snapshots during recovery plan testing. Both must be installed on the HP EVA disk arrays at the protected site and recovery site. Continuous Access and Business Copy Licensing can be checked by using HP StorageWorks Command View EVA or HP StorageWorks Storage System Scripting Utility (SSSU) Command Line Interface. Figure 7 shows the licensing information reported by Command View and SSSU.

Note: Snapclones, mirrorclones, and the ability to choose different disk group for clones are not currently supported with Site Recovery Manager or HP EVA Virtualization Adapter.

Figure 7. EVA Licensing in Command View and SSSU

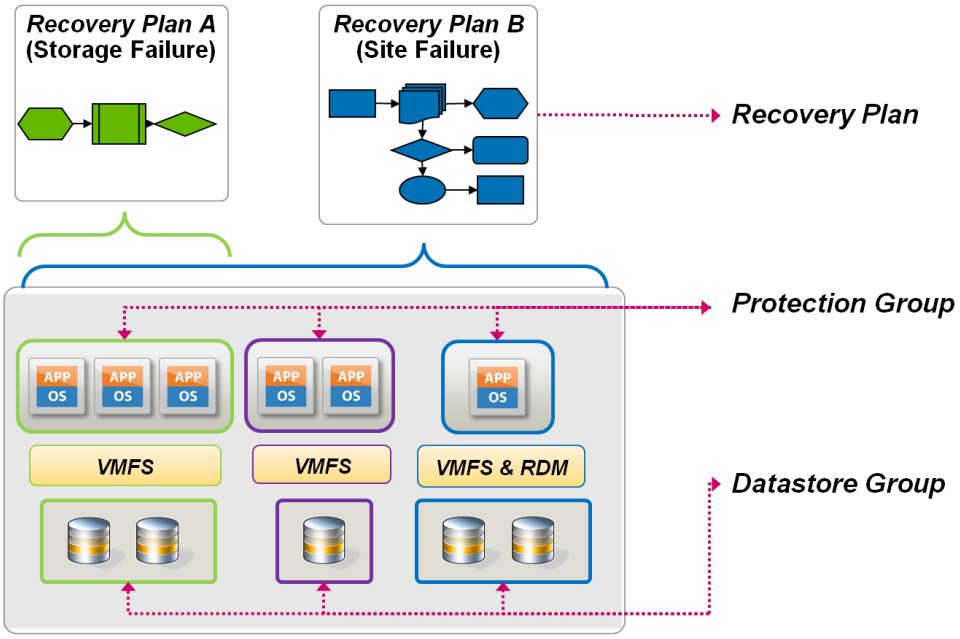




Site Recovery Manager Objects relationship

Site Recovery Manager introduces three management objects to handle the various elements of the configuration. These objects have a hierarchical relationship, as shown in Figure 8.

Figure 8. Site Recovery Manager Object hierarchy



Datastore Groups

At the lowest level, there are LUNs exposed by the EVA that are or will be replicated. In our example, we have two EVAs with five LUNs presented to the ESX Host. Each LUN houses one VMFS volume or is used as RDM. The Site Recovery Manager groups these LUNs together into datastore groups which are the smallest groupings of storage that must be failed over all together. If a VM or a VMFS volume spans multiple LUNs, the related LUNs are also placed into the same datastore group. It's important to note that Site Recovery Manager calculates this for you so you don't need to figure this out manually.

Protection Group

Each VM that is failed over is associated with a protection group, the next level in the hierarchy. A protection group is simply the set of VMs that sit on a datastore group that are failed over together. Datastore groups and protection groups have a 1:1 mapping, but the datastore group is concerned with storage, and the protection group is all about VMs. It is made up of replicated VMs and various metadata.

Recovery Plan

These previous two levels cover the objects on the protection side, but the key object on the recovery side is the recovery plan. A recovery plan is the next level in the hierarchy and is a workflow definition that applies to one or more protection groups and the corresponding instructions for recovering their VMs. Protection groups can be a part of multiple recovery plans to achieve various level of granularity. In our example, Recovery Plan A can restart a single protection group (the left most group in Figure 8), while Recovery Plan B is used to restart everything that is protected.

Recovery plans have three sections: high, medium, and low priority. The VMs in the recovery plan can be mixed and matched with the recovery sections however you choose. The high priority section starts all of its

VMs in serial order because these VMs generally have dependencies between each other, and because we want the most predictability. The medium and low priority sections start their VMs in parallel to provide the best RTO. VMs will be brought up in parallel based on the number of ESX hosts that you have. If there are four ESX hosts at the recovery site then four medium and low priority VMs can come up at once. If there is only one ESX host at the recovery site then virtual machines will be brought up one VM at a time. The number of parallel startups is limited to 16 even when there are more than 16 ESX hosts.

Configuring data replication

Understanding how Site Recovery Manager integrates with array-based replication components is key to successful deployment.

With Site Recovery Manager, the smallest possible unit of storage for data replication is a LUN. LUNs are EVA Vdisks presented to the ESX hosts. Vdisks are the unit of replication for the arrays; however, Continuous Access EVA uses a concept of consistency groups (DR group) to replicate Vdisks. Therefore, a DR group represents the smallest possible granularity for failover.

It is not possible to fail over a single Vdisk/LUN in a multi member DR group without failing over all the group members. This means that you must group virtual machines on LUNs, and you must also plan the grouping of the Vdisks into DR groups. A DR group can contain virtual disks from multiple EVA disk groups, but all DR group member virtual disks must be in the same array and use the same preferred controller on the array.

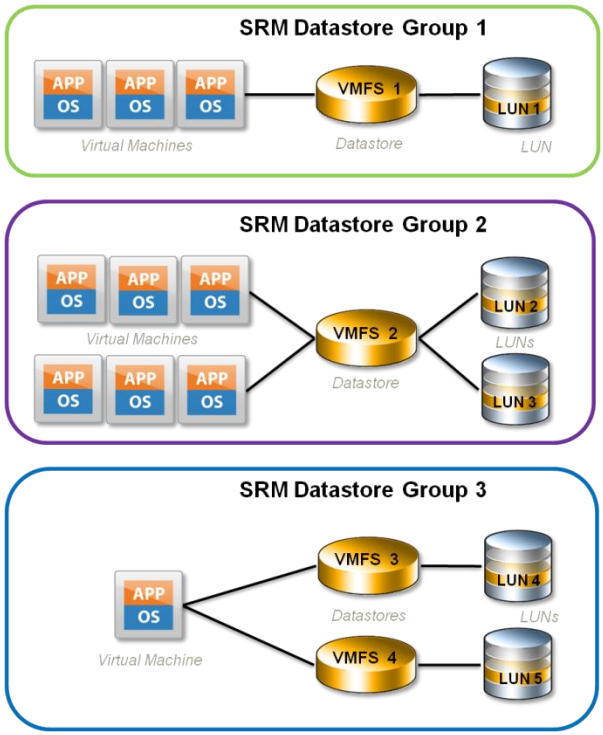
The EVA provides a limited number of DR groups[1] so understanding your environment and its replication granularity requirements can help you reduce the number of DR groups required for your environment and provide improved efficiency.

VMware formats LUNs with VMFS to store virtual machines. These VMFS -formatted LUNs are referred to as datastores. Datastores commonly contain only a single LUN but do have the ability to span across multiple LUNs. As an alternative, administrators can use raw device mapping to attach a LUN directly to the virtual machine. In this case, the LUN is formatted with the native file system of the VM. With Site Recovery Manager update 1, VMFS and Raw Device Mapping (virtual and physical) are supported for data replication.

The smallest group of datastores (and therefore LUNs) that can have its contents failed over with Site Recovery Manager is referred to as a datastore group. These groupings are calculated for you so that you do not have to worry about figuring them out. In Figure 9, LUN 1 is formatted with VMFS, and the volume is named VMFS1 and is shared by three virtual machines. It has no other dependencies and sits in its own datastore group.

[1]**Note:** Max number of DR groups may vary, depending on the EVA firmware. Consult the Continuous Access EVA release notes for the number of DR groups.

Figure 9. Site Recovery Manager datastore groups



Two factors determine what causes LUNs and datastores to be grouped together and not distinctly managed.

First, a datastore spanning multiple LUNs causes those LUNs to be grouped together in the datastore group. Failing over part of a datastore is not possible. In our example, the extended volume VMFS2 spans across LUNs 2 and 3. Since all six virtual machines on that datastore sit only on volume VMFS 2 and touch no others. VMFS 2 (and therefore LUNs 2 and 3) is alone in the second datastore group.

Second, a virtual machine can have multiple virtual disks and those virtual disks may reside in different datastores or point to a Raw Device Mapping (RDM). In that case, RDMs and datastores are forced together into a datastore group so that you do not try to fail over only part of a virtual machine. In our example, LUN4 is formatted with VMFS3, and LUN5 is formatted with VMFS4. These LUNS are grouped in a third datastore group because the virtual machine has a virtual disk in each VMFS datastore.

More than one virtual machine can, and generally should, be stored on each datastore or volume. With Site Recovery Manager, choosing where to put virtual machines should be driven by the protected applications service level agreement, the virtual machine's relationships to each other, and performance. In general, virtual machines that do not have a relationship to one another should not be mixed on the same volume nor they should be mixed in the same Continuous Access DR Group. For example, a three-tier application consisting of a database server, a web front end server, and the business logic application server should be placed into the same protection group.

If virtual machines that have no relationship are consolidated on a single volume, those virtual machines will have to failover together. With certain I/O-demanding applications, it is sometimes advisable to spread the data files onto several virtual disks using either VMFS or RDM or a combination of both. Typical examples where this approach is recommended are MS Exchange or SQL server environments. It is also recommended to isolate the database files and transaction logs in separate Volumes/LUNs. It helps to increase data resiliency in the event of a multiple drive failure and can improve performance avoiding random IOPS in the database to compete with the sequential streams of the transaction logs.

When you configure a new protection group, Site Recovery Manager requires a location to store the temporary inventory files for the protected virtual machines at the recovery site. Virtual Machine placeholders are automatically created on the recovery ESX host (The recovery host is designated in the inventory mapping). A virtual machine placeholder is a stub that represents a protected VM from the protected site Virtual Center, these are *.vmx files containing the protected VM definition files. The placeholders appear in VMware Infrastructure Client at the recovery site as any other Virtual Machine; however, you cannot modify their settings. They are used during the execution of a recovery plan when registering the recovered VMs. It is preferable to locate these temporary files on a non replicated datastore at the recovery site.

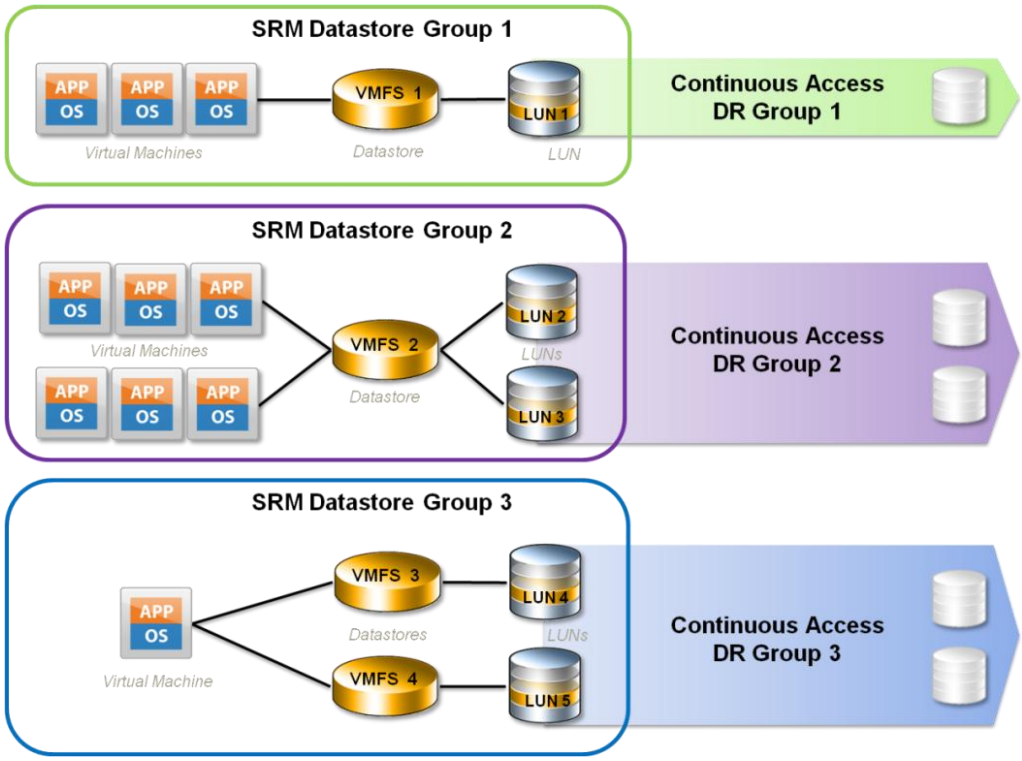
Once you have identified which virtual machines to protect, you need to configure Continuous Access to replicate the underlying LUNs and set Site Recovery Manager datastore groups for those protected virtual machines.

You should only store protected virtual machines on the datastores that are being replicated from the protected site to the recovery site. A careful review of the current VM locations is an important step in the preparation for a Site Recovery Manager installation. Most environments are not initially designed with Site Recovery Manager or Disaster recovery in mind. As a result, some mission critical VMs can possibly share the same datastore with other non critical VMS. VM placement may need to be revisited to better suit your business continuity requirements. VMware Storage VMotion can be used to relocate the virtual machines that must be protected onto Site Recovery Manager datastore groups without any interruption to services during setup.

Datastore group to Continuous Access DR group mapping

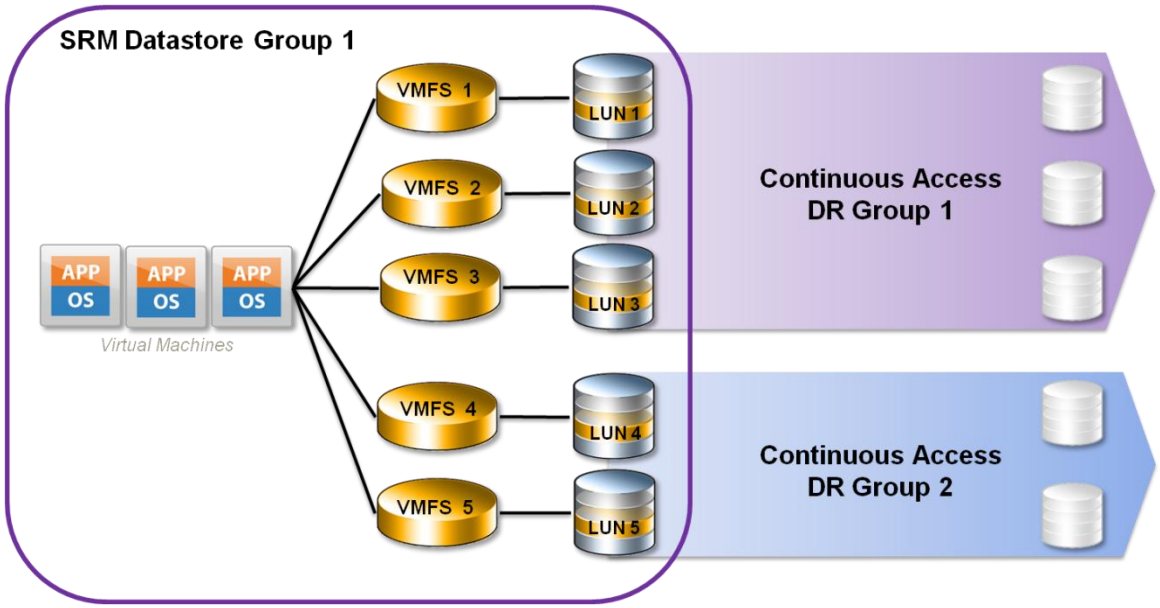
Protected datastores and RDM Vdisks can be replicated using one or more Continuous Access DR groups, but a given Vdisk can only belong to one DR group. The best practice is to establish a 1:1 relationship between the Site Recovery Manager datastore groups and Continuous Access DR groups. With straightforward mapping, the configuration is easier to plan and maintain. Figure 10 shows a symmetric datastore to DR group configuration.

Figure 10. Symmetric protection groups



Alternatively, Site Recovery Manager also allows asymmetric or one-to-many datastore group to DR group relationships, as shown in Figure 11. Although more complex to manage, you can use an asymmetric configuration to spread large protection groups with many LUNs or high capacity across multiple Continuous Access DR groups. Multiple DR groups can help to optimize the normalization time (full copy to synchronize EVA Vdisk between the source and destination array), allows replicated I/O load balancing between EVA controllers and provide additional data replication tunnel resources to replicate incoming writes. However, in most cases the same benefits can be yielded when segmenting the protected VMs and their associated storage LUNs into multiple Site Recovery Manager datastore groups each mapped to a single DR group.

Figure 11. Asymmetric datastore group configuration



Optimizing performance

ESX SAN load balancing and HP StorageWorks EVA

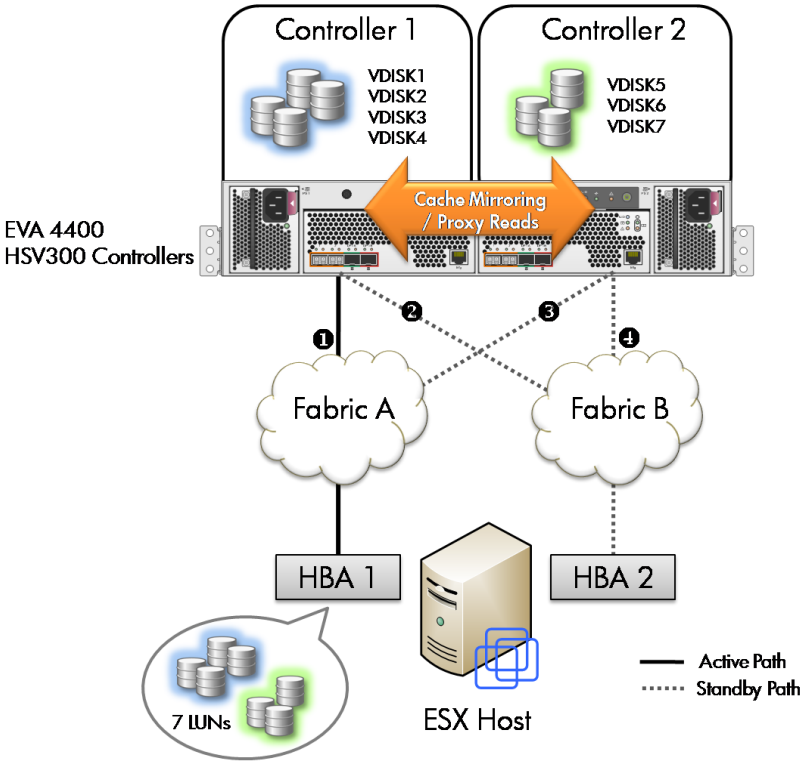
The EVA 4x, 6x, and 8x models are active/active arrays and are compliant to the Asymmetric Logical Unit Access (ALUA) industry standard for LUN failover and I/O processing. Each EVA array includes two storage controllers and allows I/O through both controllers. However, only one controller is assigned as the preferred (managing) controller of the LUN. The managing controller can issue I/O commands directly to the virtual disk. The non-managing controller, better known as the proxy controller, can accept I/O commands but cannot issue I/O commands to the Vdisk.

If a read I/O request arrives on the non-managing controller, the read request must be transferred to the managing controller for servicing. The managing controller issues the I/O request, caches the read data, and mirrors that data to the cache on the non-managing controller, which then transfers the read data to the host. Because this type of transaction, called a proxy read, requires additional overhead, it potentially provides less than optimal performance. (There is little impact on a write request because all writes are mirrored in both controllers' caches for fault protection.)

When configuring VMware ESX server storage adapters and multipathing, ALUA causes additional design considerations. By default, VMware Infrastructure 3 only uses one path from the host to a given LUN at any given time; regardless of path policy. If the path fails, the ESX Server detects and selects another path for failover. A path is defined as a connection between a host bus adapter port and an EVA controller port. Using dual ported c-class blade FC HBA with the EVA4400 results in four possible paths

As shown in Figure 12, the default active path uses FC HBA1 (vmhba1) and the EVA controller port with the lowest PWWN, typically Controller 1 Port 1. The VMware ESX 3.5 multipathing stack does not automatically balance the load across all of the available data path or privilege direct path to the EVA virtual disk managing controller.

Figure 12. ESX multipathing default configuration



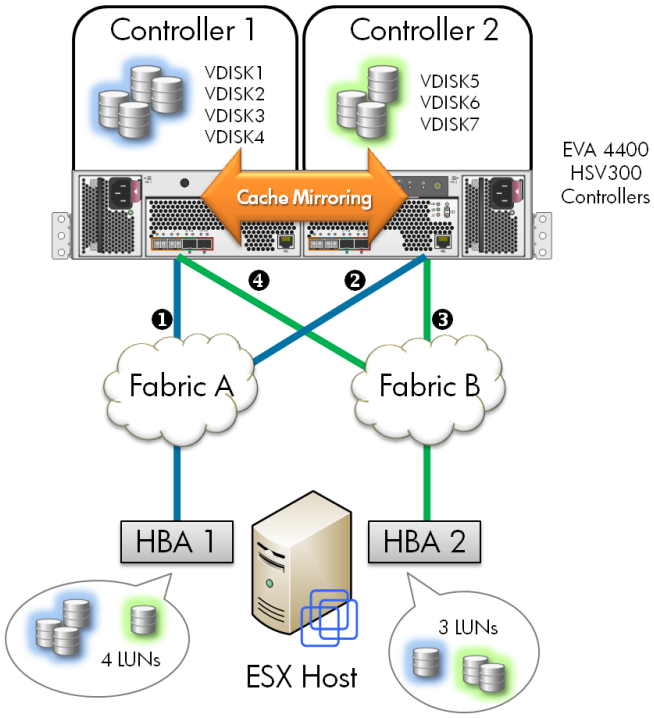
The default ESX 3.5 multipath configuration may result in imbalanced SAN and EVA controller load. This can affect performance under heavy workloads, with all I/Os being processed by only one Host Bus Adapter and one EVA controller.

To optimize storage paths performance, HP recommends configuring a preferred path on the ESX host for each LUN. Ensure the multipath policy is set to Fixed, and then manually assign a preferred path connecting to the EVA managing controller for each of the presented LUNs. HP recommends setting an identical multipathing configuration on all ESX hosts sharing storage LUNs so that, regardless of which host a VM is executing on, storage is configured and optimized the same way.

Storage Paths can be configured using VMware Infrastructure Client (Configuration Tab and then Storage Adapters), by using the command line tool `esxcfg-mpath` at the ESX server console or by using the VI toolkit (PowerShell). A scripted approach helps to consistently set multipathing in environments comprising of a large number of LUNs and ESX hosts.

Figure 13 shows an optimized connection diagram in which each LUN is assigned a preferred path, using vmhba1 for the first four LUNs and vmhba2 for the remainder.

Figure 13. Optimized ESX multipathing configuration with HP StorageWorks EVA



Implicit LUN transition

To optimize the host data paths and reduce I/O on the mirror ports, the EVA can automatically transfer the management of a virtual disk to the array controller that receives the most read requests for that virtual disk. This process is known as the implicit LUN transition. There are several Perl and PowerShell scripts available on the Web to automate ESX path assignments to leverage this behavior. For a PowerShell script example, see the VM junkie Blog. Once set on the ESX host no further configuration is required. (For example, setting the virtual disk preferred controller with Command View EVA is not required for non-replicated LUNs). The ESX multipath setting takes precedence. The Vdisks automatically switch over their managing controller on the EVA to use the most optimized path.

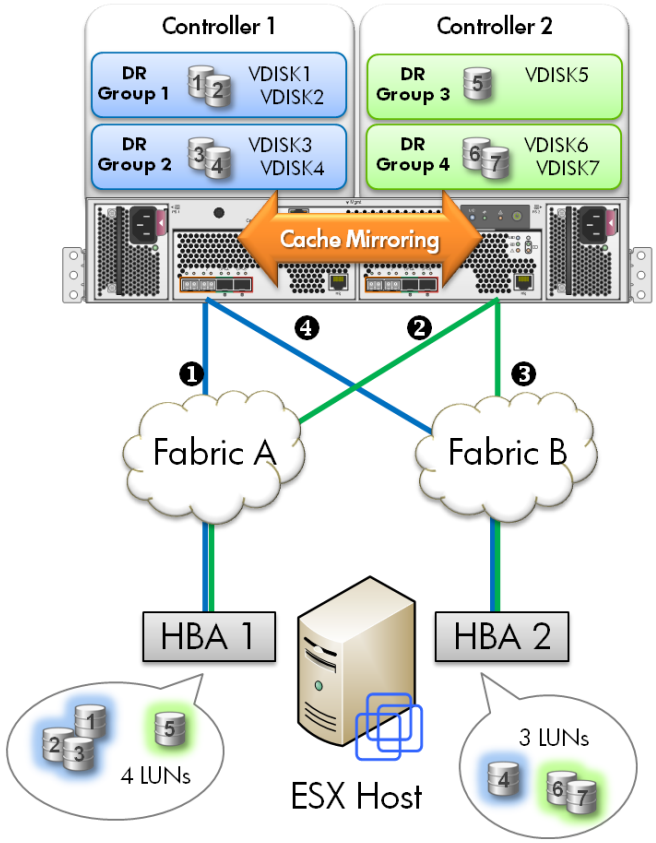
Implicit LUN transition and Continuous Access

When using Continuous Access EVA to replicate EVA virtual disks, the implicit LUN transition is automatically disabled for all members of an HP Continuous Access EVA DR group. Because Continuous Access requires that all members of a DR group be managed by the same controller, it would be necessary to move all members of the DR group if excessive proxy reads were detected on any member. This would impact performance and create a proxy read situation for the other virtual disks in the DR group which in turn may trigger another transition eventually leading to a path trashing scenario.

This is an important consideration when setting your ESX server multipath configuration. Without the support of implicit LUN transition for the replicated Vdisks, incorrect ESX host settings may cause one or more virtual disks in a DR group to experience excessive proxy reads. With this in mind, load balancing scripts should be used carefully as they may establish a less than optimal configuration, causing unnecessary proxy reads.

When using Continuous Access EVA in a VMware ESX environment, HP recommends setting a preferred controller to virtual disks configured in a DR group and reflecting the configured affinity on the ESX multipathing configuration. As shown in Figure 14, you can still balance the LUN across multiple HBA ports, but you must ensure that the destination address points to a port on the owning controller to avoid excessive proxy reads.

Figure 14. Optimized ESX multipathing with HP Continuous Access EVA



LUNs 1,2,3 use Path #1

LUN 5 uses Path #2

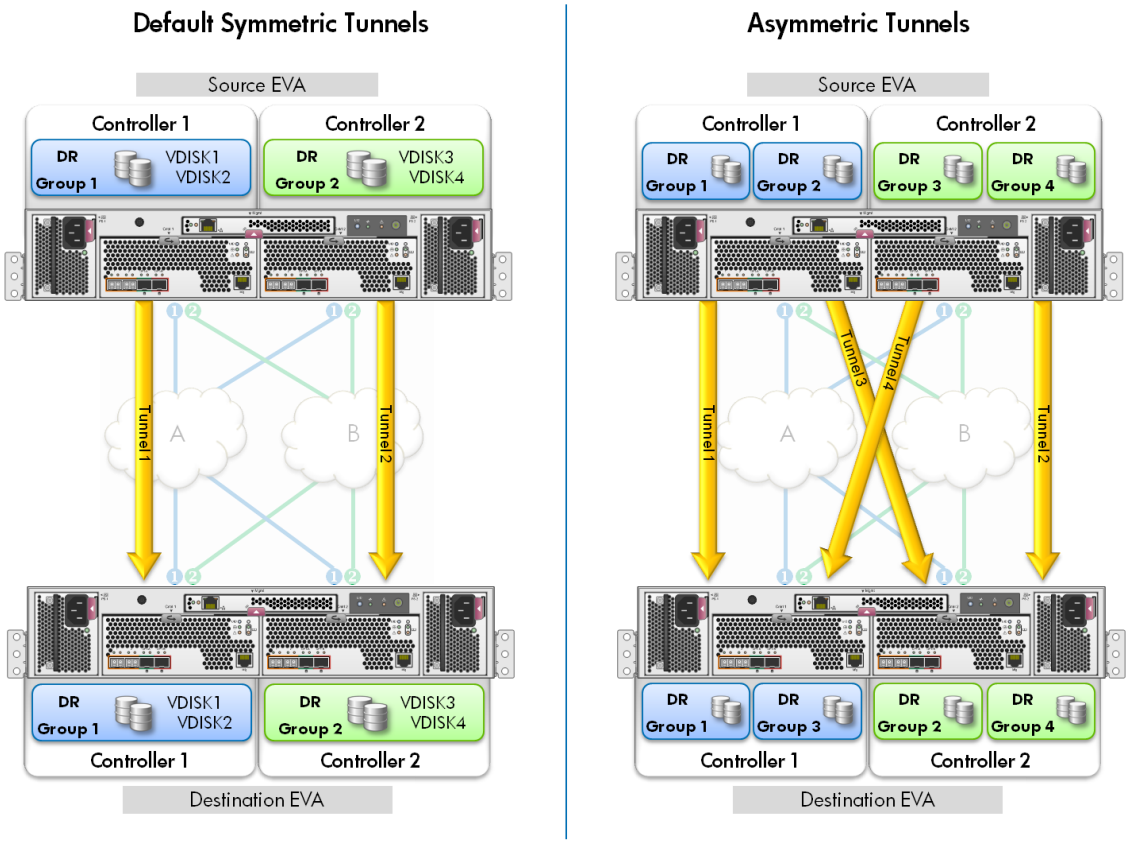
LUN 4 uses Path #4

LUNs 6,7 use Path #3

Continuous Access replication tunnels

Replicated EVA Vdisks (copy sets) are grouped in DR groups. When a DR group is created, a relationship is established between a source and destination-managing EVA controller. Data is replicated from the source to the specified destination controller through a path. This path is called a data replication tunnel (DRT). There is one tunnel per source/destination controller pair. To ensure I/O consistency, Continuous Access uses only one tunnel per DR group. These resources are limited; therefore, it is important to configure data replication adequately to scale the number of tunnels to avoid possible resource contention with large DR groups under heavy write traffic. As shown in Figure 15, with two or more DR groups, the EVA pair establishes a symmetric relationship by default (using the same controller, A or B on source and destination EVAs); which limits the number of concurrent tunnels to two.

Figure 15. Symmetric versus Asymmetric tunnels configuration



However, using an asymmetric configuration whereby a DR group is using a different controller on the source and destination EVA (for example, DR group source members are online to EVA-1/Controller A and destination members are online to EVA-2/Controller B), you can establish up to four concurrent tunnels between a source and destination EVA.

Host port preferences are set to defaults during controller initialization but can be changed using HP Command View EVA 9.0 and later. Manually setting host port preferences should be done carefully and only when absolutely necessary. Default settings are appropriate for most use cases. For more information see the HP StorageWorks Continuous Access EVA Implementation Guide.

For best performance of remote replication, the average workload (reads and writes) should be equal on both controllers and on both fabrics and ISLs. To obtain this balance, take measurements and keep the utilization rate of each ISL below 40%. If one link fails, the average utilization of the operating link should not exceed 80%. Similarly, the utilization rate of a single controller as measured on all host ports should not exceed 45% on average, or peak above 50%, to prevent overloading the surviving controller should one controller fail.

Remember, all members of a DR group must share a preferred controller. Therefore, load balancing a single application with all of its virtual disks in a single DR group (as required) across multiple controllers is not possible.

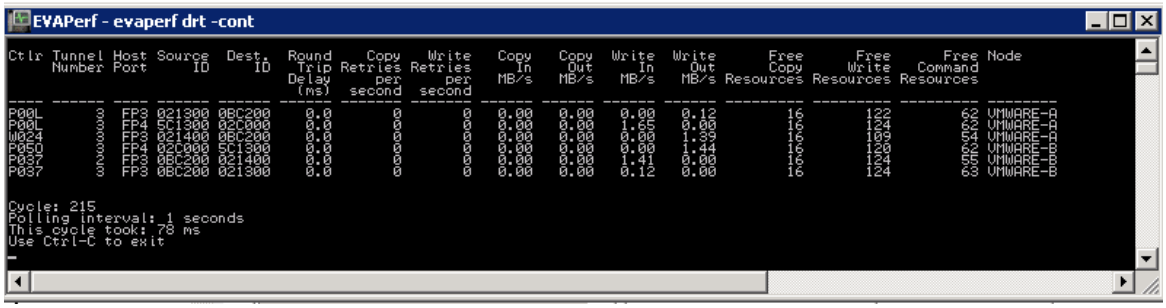
Monitoring EVA activity

You can use HP StorageWorks Command View EVA Performance Monitor (EVAPerf) to monitor and display EVA performance metrics in real time from a command line interface or a graphical user interface (Windows PerfMon). You can also display performance metrics in an external application, such as Microsoft Excel using a CSV (comma-separated value) or TSV (tab-separated value) format. EVAPerf is automatically installed by HP Command View software suite. The EVAPerf components are installed in the following directory:

```
C:\Program Files\Hewlett-Packard\EVA Performance Monitor
```

For example, the command `evaperf.exe drt -cont` shows the DR Tunnels activity. Figure 16 shows the command output. For best performance of remote replication, free write and free command resources should be available. If values for these metrics are consistently low, the DR group tunnel becomes a bottleneck. Balance the DR groups between controllers on source and destination array to create new tunnels. Up to four concurrent tunnels can be used between the arrays.

Figure 16. Command output


EVAPerf - evaperf drt -cont
Ctlr Tunnel Number Host Port Source ID Dest. ID Round Trip Delay (ms) Copy Retries per second Write Retries per second Copy In MB/s Copy Out MB/s Write In MB/s Write Out MB/s Free Copy Resources Free Write Resources Free Command Resources Node
P00L 3 FP3 021300 0BC200 0.0 0 0 0.00 0.00 0.00 0.12 16 122 62 VMWARE-A
P00L 3 FP4 5C1300 02C000 0.0 0 0 0.00 0.00 1.65 0.00 16 124 62 VMWARE-A
W024 3 FP3 021400 0BC200 0.0 0 0 0.00 0.00 0.00 1.39 16 109 54 VMWARE-A
P050 3 FP4 02C000 5C1300 0.0 0 0 0.00 0.00 0.00 1.44 16 120 62 VMWARE-B
P037 3 FP3 0BC200 021400 0.0 0 0 0.00 0.00 1.41 0.00 16 124 55 VMWARE-B
P037 3 FP3 0BC200 021300 0.0 0 0 0.00 0.00 0.12 0.00 16 124 63 VMWARE-B
Cycle: 215
Polling interval: 1 seconds
This cycle took: 78 ms
Use Ctrl-C to exit

For more information on EVAPerf use, see the Performance analysis of the HP StorageWorks EVA using Command View EVAPerf software tool white paper.

Sizing the SAN ISL

The location of the remote site often determines the ISL technologies that meet your distance and performance requirements. The following factors affect the necessary bandwidth:

- Distance—The transmission distance supported by specific link technologies
- Recovery point objective—The acceptable difference between local and remote copies of data
- Bandwidth capacity and peak loads—The effect of application peak loads on bandwidth

For the later, determining the peak load of each protected applications is a rather complex and lengthy process in large environments. Standard performance monitoring tools, such as Microsoft Perfmon, VMware ESXTop or HP EVAPerf, can be used to perform these assessments. The principle is to collect hosts, disk arrays, and applications performance metrics over a period of time. The collected metrics must then be consolidated and computed to grasp the average and near peak values for the total applications disk write throughput and finally derive the adequate bandwidth for the ISLs.

To simplify and automate remote replication link bandwidth sizing, you can use HP StorageWorks Essentials Data Replication Designer (DRD). DRD is a free GUI-based tool that simplifies collection of application profile data from customer application servers and automatically calculates the size of the link required for remote replication to meet business requirements. This tool is available for approved HP partners. For more information about Data Replication Designer, consult your HP representative or Storage Architect.

Testing

The testing is aimed at validating the configuration of a disaster tolerant solution, using bidirectional data replication for a mix of applications and workloads that include the following:

- Microsoft Exchange Server 2007 in a Windows Server 2008 environment
- Microsoft SQL server 2008 in a Windows server 2008 environment
- Microsoft Internet Information Service (IIS)
- File server in a Windows server 2003/Windows server 2008 environment
- My SQL *Plus Web front end in a Red Hat Linux 5 environment

Site Recovery Manager is configured with four protection groups protecting 13 virtual machines. Replicated storage was a mix of Raw Mapped LUNs and VMFS datastores.

Protection Group 1

- Exchange 2007 (1500 mailboxes)
- Three Protected Virtual Machines: Mailbox Server, Client Access, and Hub Transport
- Data replication from Site A to Site B

Virtual Machine Disks	Type	Purpose	LUN	DR group
CAS C:\(40 GB)	VMFS	OS files and application binaries	1	DR Group 1
Hub C:\(60 GB)				
Mailbox C:\(40 GB)				
Mailbox E:\(1 TB)	RDM	Exchange Storage Groups 1-7	2	
Mailbox F:\(1 TB)	RDM	Exchange Storage Groups 8-14	3	
Mailbox E:\Logs (140 GB)	RDM	Exchange Logs 1-7	4	
Mailbox F:\Logs (140 GB)	RDM	Exchange Logs 8-14	5	

Protection Group 2

- Microsoft SQL Server 2008 plus IIS
- 3 Protected Virtual Machines: SQL Server 2008, IIS and a Windows 2008 application client
- Data replication from Site B to Site A

Virtual Machine Disks	Type	Purpose	LUN	DR group
SQL Server C:\ (40 GB)	VMFS	OS files, application binaries, and static html content	6	DR Group 2
IIS Server C:\ (40 GB)				
Test Client C:\ (60 GB)				
SQL E:\(400 GB)	VMFS	SQL databases	7	
SQL F:\(150 GB)	VMFS	SQL transaction logs	8	

Protection Group 3

- Three protected Windows 2008 File Servers
- Data replicated from Site B to Site A

Virtual Machine Disks	Type	Purpose	LUN	EVA DR group
Server 1 C:\(40 GB)	VMFS	OS files	9	DR Group 3
Server 2 C:\(40 GB)				
Server 3 C:\(40 GB)				
Server 1 D:\(800 GB)	VMFS	User data	10	
Server 2 D:\(400 GB)	VMFS			
Server 3 D:\(500 GB)	RDM	User data (migrated from Physical Server)	11	

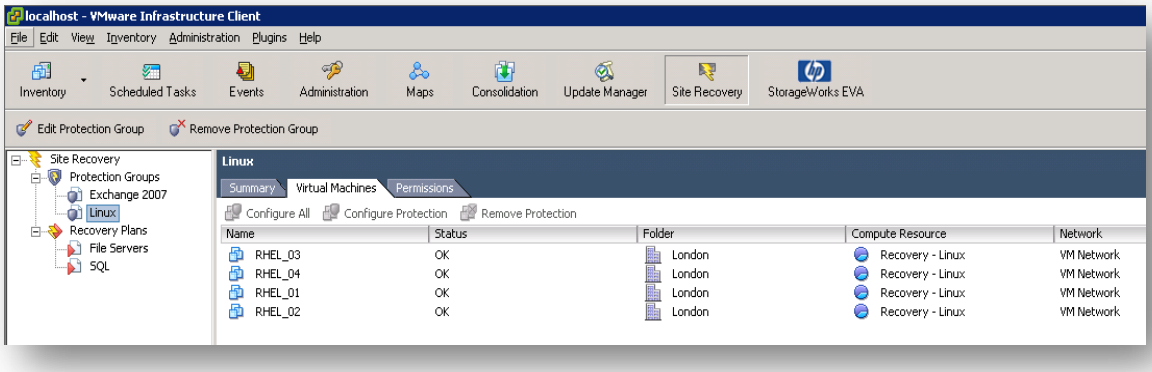
Protection Group 4

- 4 protected Virtual Machines, Red Hat Linux Enterprise 5
- Data replicated from Site A to Site B

Virtual Machine Disks	Type	Purpose	LUN	EVA DR group
RHEL 1 Boot	VMFS	OS files and application binaries	12	DR Group 4
RHEL 2 Boot				
RHEL 3 Boot				
RHEL 4 Boot				
Data 1	VMFS	My SQL DB and PHP	13	
Data 2	VMFS	My SQL transaction logs	14	
Data 3		Apache data		

In our environment, MS Exchange 2007 and Linux virtual machines run at Site A and replicate to Site B. MS SQL Server 2008, IIS and the file server VMs run at Site B and replicate to Site A Figure 17 shows the Site Recovery Manager configuration summary for Site A with the local Exchange and Linux protection groups and the recovery plans for SQL Server and File Servers from site B.

Figure 17. Site A protection groups and Site B recovery plans



Performance testing

This section of the testing aims to characterize the performance overhead attributable to synchronous data replication.

Microsoft Exchange 2007

We used Microsoft JetStress to exercise the Mailbox server storage. The environment is configured with 14 Exchange storage groups. The JetStress utility identifies the maximum I/O throughput that a given configuration can sustain. Each test run is set for a period of two hours. To baseline the throughput, the same JetStress tests are executed with and without data replication.

The disk configuration for the protected and non protected virtual machines is identical. Each configuration is presented with the same set of Exchange logical unit numbers (LUNs) from the EVA array that uses two disk groups. There is no physical disk sharing between the two LUNs housing the Exchange databases and the two LUNs housing the Exchange transaction logs. All EVA LUNs are configured as RAID 1. The virtual machine boot disk is housed on a shared VMFS volume, while the Exchange data disks use Raw Device Mapping. Table 2 shows the disk layout.

Table 2. Exchange disk layout

Virtual Machine Disk	Purpose	LUN
C:\(VMFS)	OS files and application binaries	1
E:\(RDM)	Exchange storage groups 1-7	2
F:\(RDM)	Exchange storage groups 8-14	3
E:\Logs (RDM)	Exchange logs 1-7	4
F:\Logs (RDM)	Exchange logs 8-14	5

In previous testing we established that VMFS and Raw Device Mapping perform equally well when subjected to an Exchange 2007 workload. In this configuration, we configure the Exchange storage with RDM to leverage the Microsoft Exchange database transportability feature. This Exchange 2007 feature can easily migrate mailbox databases between servers (physical or virtual) when using a SAN. See the HP Best practices for deploying an Exchange 2007 environment using VMware ESX 3.5 white paper for recommendations on configuring Microsoft Exchange 2007 in a VMware ESX environment.

The data in Table 3 compares the differences in I/O performance (response time and I/O request rate) between the physical server and the VM by using VMFS and RDM storage access management for all of the logical disks in the two-hour test period.

Table 3. Differences in I/O performance

Database I/O (Vraid 1)	**Non-protected VM**	**Protected VM (One DR group)**	**Protected VM (Two DR groups)**
Average database disk transfers/second	10,454.432	9,671.55	9,752.81
Average database disk reads/second	5,592	5216	5,352
Average database disk writes/second	4,862	4454	4,399
Average database disk read latency (ms)	0.014	0.014	0.014
Average database disk write latency (ms)	0.008	0.010	0.009
Log I/O (Vraid 1)	**Non-protected VM**	**Protected VM (One DR group)**	**Protected VM (Two DR groups)**
Average database disk writes/second	2,926	2,426	2,480
Average database disk write latency (ms)	0.001	0.001	0.001

The JetStress results indicate that the protected VMs provide comparable storage performance to that of a non-protected VM when replicating data over a low latency FC inter site link (<1ms). Data replication affects the application I/O performance in our test configuration by less than 10% due to a slight increase of the write response times. Using one or two Continuous Access DR group(s) to replicate the five virtual disks provide similar I/O throughput for the workload tested. However, spreading the EVA virtual disks into two or more DR groups can optimize full copy or merge operations between the source and destination arrays.

DR group full copy performance

We measure and compare the DR group normalization (full copy) performance using one and two DR group(s) to mirror the Exchange Data set. The configuration comprises five virtual disks for a total net capacity of 2.2 TB. All Vdisks are configured with Vraid 1.

Configuration 1 uses a single DR group to replicate all five Vdisks, Configuration 2 uses two DR groups to replicate the five Vdisks. The charts in Figure 18 and Figure 19 show the measured throughput during the normalization window.

Figure 18. Full copy using a single DR group

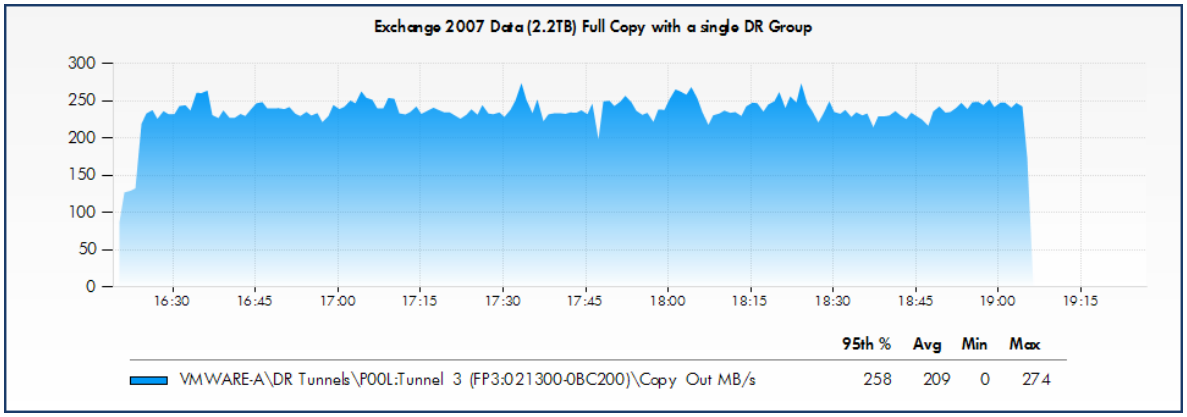
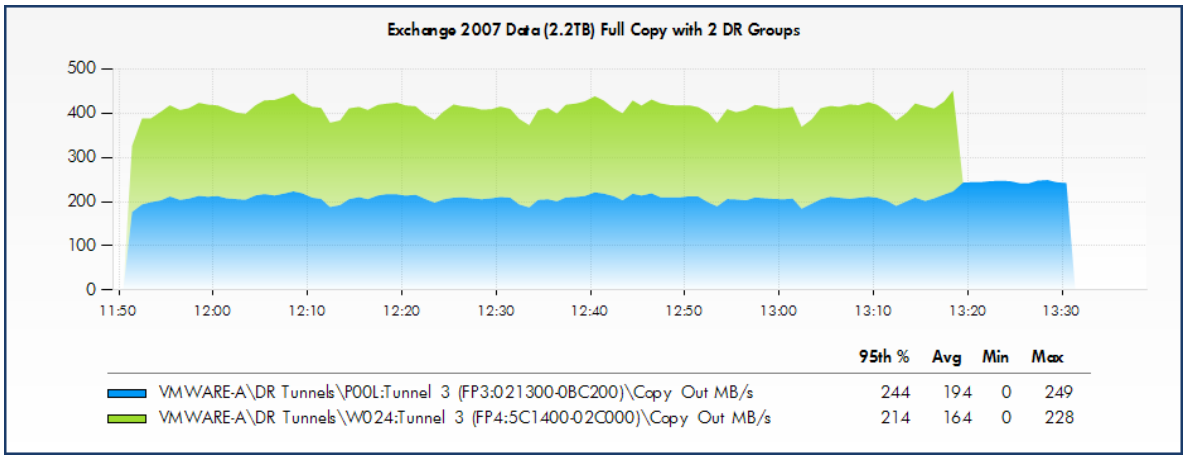


Figure 19. Full copy using two DR Groups



As shown in Table 4, spreading the application data over two DR groups yielded a 65% performance improvement, shaving approx one hour off the normalization window. You can configure Site Recovery Manager protection groups to use multiple Continuous Access DR groups.

Table 4. DR group normalization time

Configuration	Normalization time	Average throughput
1 x DR group	2 hours, 46 minutes	226 MB/s
2 x DR groups	1 hour, 40 minutes	375 MB/s

Failover considerations

Failover occurs when a Site Recovery Manager recovery plan is executed, and the Site Recovery Manager recovery plan is configured to fail over Site Recovery Manager protection groups that use replicated EVA Vdisks as VMFS datastores or Raw Mapped LUNs. During normal operation, the replication path is from a protected site (local site) to a recovery site (remote site). When a failure occurs at the local site (due to hardware failure or the entire site loss) an administrator can initiate a failover. The replication path is reversed from the recovery site to the protected site and the virtual machines are restarted at the recovery site. With HP EVA Replication Adapter and Site Recovery Manager, the failover process is an automated feature and the process is executed, using a recovery plan defined at the recovery site.

Site Recovery Manager also enhances the efficiency of testing disaster recovery plans. In this case, failover executes in an isolated environment, and the environment is quickly and automatically cleaned up following the testing to avoid disrupting operations. As a result, the IT team can minimize impact on the production environment and avoid downtime for customers. Site Recovery Manager allows administrators to initiate a test failover every time the environment goes through significant modifications or reconfiguration that can possibly affect the recovery of the protected virtual machines and applications.

In a test failover, data replication is not suspended and the protected virtual machines keep running. To simulate a storage unit failover, a snapshot is created for each of the replicated EVA Vdisk that is used by the protected virtual machines. The snapshots are exposed to the recovery host and the protected VMs are started in an isolated network environment. This is a nondisruptive process, but a heavily loaded array may experience slight performance degradation when the snapshots are established. In addition, test failover in some cases causes another side effect that may require additional planning prior to executing a test failover. The non disruptive nature of the test failovers simulates an unplanned failover where the production VMs are running in parallel with the recovered VMs.

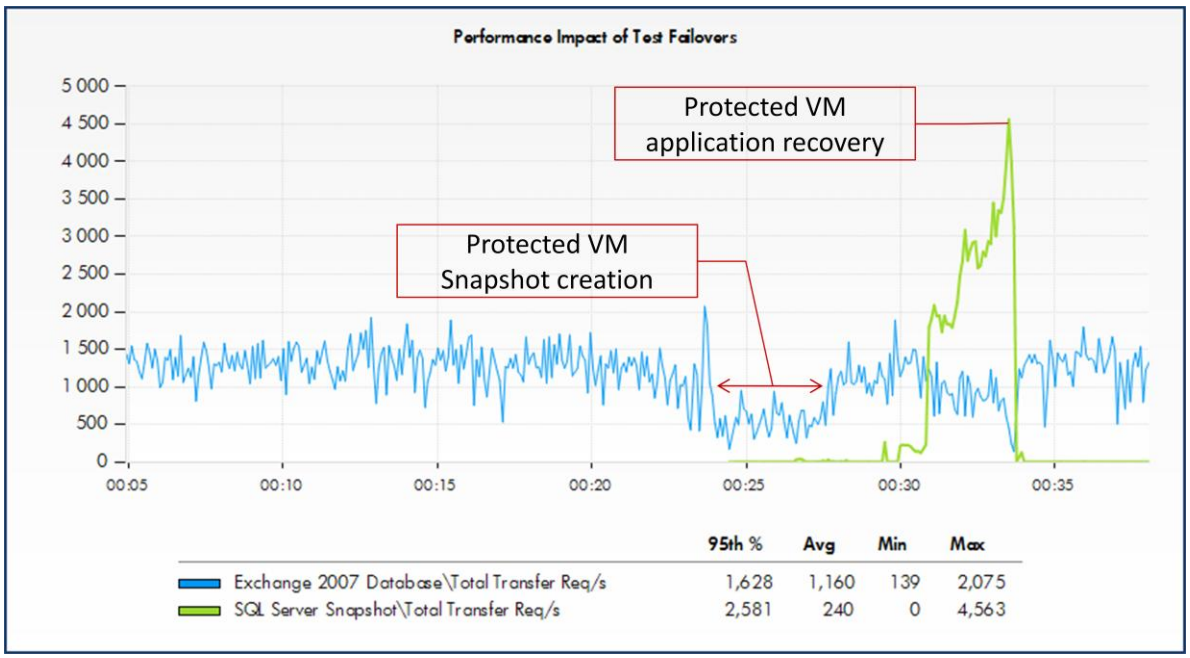
Because databases are not quiesced during an unplanned failover, a number of transactions that were not flushed to the database files exist. As a result, a consistency check or a soft recovery is usually required when the VM and application are started during the execution of a test failover. This primarily affects transactional applications, such as MS Exchange, SQL server, or My SQL.

Note: Soft recovery is used as a generic term here and refers to the action of rolling forward or backward the transaction log to bring an application database to a consistent and up-to-date state.

Failover in test mode provides an accurate unplanned failover simulation, including the application recovery intricacies. However, transaction log replay and database consistency checks can generate a significant amount of additional IOPS on the snapshots, which may affect the performance of the replicated Vdisks parents and the housed applications. For this reason, HP recommends that you avoid peak production hours to execute test failovers in environments where both sites are hosting production servers.

The impact of test failovers that include application recovery can be mitigated with adequate disk array sizing. These additional IOPS must be factored in when sizing the EVA disk group to allow transparent failover testing during production hours. When deploying Site Recovery Manager in an active/passive datacenter scenario, the impact of test failover on I/O performance is negligible. Figure 20 shows a slowdown of an Exchange 2007 mailbox server running at site A, caused by the test of a recovery plan for a SQL server application located at Site B.

Figure 20. Limited performance impact caused by a test failover



In our test environment, the performance impact is limited in both intensity and duration. This is because there is sufficient headroom on the array as we include some margin in the sizing exercise to support application recovery testing and accommodate for future growth. However, smaller configurations with less physical drives can exhibit more important disruption.

Networking consideration

Some applications have a strong dependency on infrastructure components, such as DHCP, DNS, or Active Directory domain controllers (DC). These components often use different protection techniques to provide resiliency and may not be part of a Site Recovery Manager recovery plan. This can preclude a successful execution of the plan from an application perspective, especially when using the *bubble* virtual network. For example, failing over protected Exchange 2007 mailbox servers requires domain controller availability. If DC is not available in the isolated network environment, the mailbox servers will not start successfully and cannot mount the mailbox databases.

Recovery plan and failover considerations

Once started, recovery plans have minimal interaction capabilities. Therefore, it is important to use Site Recovery Manager test functionality to practice and validate the plan prior to execution. Doing so ensures proper startup order of VMs and application dependencies. You can (and should) create multiple recovery plans that are based on application tiering and dependency. Multiple recovery plans can be started simultaneously to bring up the environment faster after a disaster occurs.

Failback considerations

Failback is the process of setting the replication environment back to its original state at the protected site prior to failover. Configuring the failback process with Site Recovery Manager is a manual process that occurs after failover and the steps vary with respect to the degree of failure at the protected site. For example, the failover could have been due to an array failure or the loss of the entire data center. The manual configuration of failback is important since the protected site could have completely different hardware and/or SAN configuration after a disaster occurred. Using Site Recovery Manager, once failback is configured it can be managed and automated like any planned Site Recovery Manager failover. The recovery steps can differ based on the conditions of the last failover that occurred. If failback follows an unplanned failover, a full data re-mirroring between the two sites may be required. This step usually takes the most time in a failback scenario.

Failback scenario

The following failback scenario uses Site Recovery Manager as a failback mechanism to Site A after executing at Site B, with recovery plan R1 in recovery mode. Follow these steps to copy the scenario used in testing.

Note: If you are preparing failback after a planned failover, this step can be skipped.

1. Establish appropriate array replication from Site B to Site A with Command View EVA. Include all Vdisks assigned to the protected virtual machines (VMFS datastores and Raw Device Mapping).
    a. Create or reconfigure DR groups between Site B and Site A, with Site B virtual disks as the source and Site A as the destination.
    b. Set host presentation for all target Vdisks to ESX host(s) at Site A.
    c. Set DR group destination host access to **None.**

Using VMware Infrastructure Client with Site Recovery Manager Plugin

2. Delete recovery plan R1 at Site B.
3. If Site A still has protection group or groups P1 in recovery plan R1, delete protection group(s) P1 at Site A.
4. If it is not already in place, use the VI Client at Site B to establish Site A as the secondary site for Site B. (If Site A does not have Site Recovery Manager installed, install Site Recovery Manager and the HP EVA Replication Adapter at Site A.)
5. Create protection group or groups P2 at Site B to add virtual machines that must failback to Site A.
6. At Site A, create a recovery plan R2 for the protection group or groups P2.
7. After the user has determined with Command View that the virtual machines Vdisks have been fully replicated to Site A, execute recovery plan R2 at Site A in test mode.
8. If the test is successful, execute recovery plan R2 in recovery mode.
9. At this point, you may want to re-protect recovered virtual machines to Site B:
10. Delete recovery plan R2 at Site A.
11. Delete protection group(s) P2 at Site B.
12. Create protection group(s) P3 at Site A to protect recovered virtual machines
13. At Site B, create a recovery plan R3 for the protection group or groups P3.

Best practices and caveats

This section includes recommendations for customers who plan to deploy VMware vCenter Site Recovery Manager, using HP StorageWorks Continuous Access EVA. The following best practices and recommendations should be considered by system administrators to successfully deploy Site Recovery Manager in their virtual environment.

ESX server and Site Recovery Manager administration

- When storage is configured with VMFS on ESX 3i, you cannot dynamically resize an existing VMFS partition. However, you can increase the volume size by using an extent to tag a new partition to an existing VMFS datastore. HP does not recommend using VMFS extents because they tend to make the storage replication configuration harder to manage. Instead, when running out of space on a VMFS datastore, HP recommends creating a new LUN/VMFS data store and moving some data onto it. Consider using Storage VMotion to avoid service interruption.
- Do not share the VMFS datastores housing I/O demanding application, such as application databases and logs, with other virtualized applications. However, you can share a VMFS datastore to house VM files (configuration, vmdk, and so forth).
- Configure multipathing with the FIXED policy and assign a preferred path for each LUN presented to the ESX host. Balance LUNs across all available FC HBA to maintain an equal workload on across EVA controllers and Fabric ISLs.
- When possible, only store protected virtual machines on the datastores that are being replicated from the protected site to the recovery site. A careful review of the current VM locations is an important step in the preparation for an Site Recovery Manager installation
- Protected datastores and RDM Vdisks can be replicated using one or more Continuous Access DR groups. However, it is preferable to establish a 1:1 relationship between the Site Recovery Manager datastore groups and Continuous Access DR groups. With straightforward mapping, the configuration is easier to plan and maintain. The exception to the rule is a protection group, combining a single Virtual Machine with multiple dataset spread over several virtual disks (for example, MS Exchange 2007 or SQL server 2008). In this case, VMFS or RDM disks can be spread into two or more Continuous Access DR groups to optimize full copy and data resynchronization between source and destination arrays, using two or more replication tunnels.
- Site Recovery Manager recovery plans executed in test mode simulate unplanned failover situation from an application perspective. Increased disk activity may result from database application recovery activity on the tested Virtual Machines. Size the recovery side array adequately to withstand the I/O increase or defer test activities to take place during off peak hours.

Storage administration

- EVA disk group layout: To optimize performance and provide fault tolerance, place application databases files, transaction logs, rich media content, and ESX VM files on separate physical disks in different disk groups. This enables recovery of data to the point of failure. If the log files and databases are maintained on the same disk group, in the unlikely event of a disk group failure, recovery to the point of the last known good backup is the only possibility. Performance also improves because database I/O operations are random, whereas transaction log and rich media file operations are sequential. Having both on the same physical disk can cause I/O performance degradation.
- Implicit LUN transition is disabled for EVA Vdisks configured for data replication. HP recommends balancing the Continuous Access DR group across EVA controllers to maintain an equal workload and to use concurrent replication tunnels. Set the Vdisks PREFRRED_PATH by using Command View EVA. Multipath settings on the ESX host must be set accordingly to avoid excessive proxy reads.
- To simplify and automate remote replication link bandwidth sizing, use HP StorageWorks Essentials Data Replication Designer. Data Replication Designer (DRD) is a free GUI-based tool that simplifies collection of application profile data from customer application servers and automatically calculates the size of the link required for remote replication that meets the business requirements. This tool is available for approved HP partners. For more information about Data Replication Designer, consult your HP representative or Storage Architect.

Implementation caveats

Renamed Vdisks on the EVA

The EVA SRA requires the source and destination Vdisk to have the same name. Command View EVA allows renaming of either the source or destination Vdisk in a copy set. However, if the virtual disk name is different on source and destination array, Site Recovery Manager will not correctly identify the protected disk and will fail to create the snapshot when executing a recovery plan in test mode.

HBA rescan problem

In our testing, we sometimes observed that QLogic HBAs needed to be rescanned twice before the EVA LUNs (VMFS volumes or raw-mapped LUNs) are discovered. When using Site Recovery Manager, the HBA rescan only occurs once by default. The default Site Recovery Manager rescan behavior can be changed in the vmware-dr.xml file located in the c:\Program Files\VMware\VMware Site Recovery Manager\config directory.

For more information, see the VMware KB 1008283.

Protection group to DR group relationship

There is no mechanism in Site Recovery Manager that prevents an administrator from creating two different protection groups that use Vdisks contained within the same Continuous Access DR group. Configuring a protection group this way could have dire consequences.

Conclusion

The test results and related information contained in this paper clearly illustrate how to properly plan for and successfully deploy a fully operational remote copy infrastructure, using VMware vCenter Site Recovery Manager in conjunction with HP Continuous Access EVA and the EVA VMware Site Recovery Adapter.

Key planning and deployment considerations include the following:

- VMware vCenter Site Recovery Manager controls and orchestrates failover and test operations by using a software component called the Storage Replication Adapter (SRA). The EVA VMware Site Recovery Adapter (EVA SRA) provides the communication layer between VMware SRA and the EVA array. The EVA SRA must be installed on both the local (protected) server and the remote (recovery) Site Recovery Manager servers.
- It is important to understand the hierarchical relationships between VMware recovery plans, VMware protection groups, and VMware datastore groups, as well as how each maps to EVA DR groups. In general, Site Recovery Manager datastore groups should map 1-1 with EVA DR groups, though it is possible under certain circumstances for a datastore group to map to multiple EVA DR groups.
- The addition of Continuous Access EVA, when used in conjunction with VMware, significantly changes the planning considerations for implementing multipathing between the VMs and the EVA array.
- When upgrading from a standalone to a protected environment with Site Recovery Manager, virtual machine physical placement may have to be re-evaluated and modified to ensure that all VMs and their data elements are appropriately mapped into the described hierarchies.

Key operational considerations include the following:

- Site Recovery Manager provides the ability to fully test a disaster recovery plan by executing a failover in an isolated environment. This environment is quickly and automatically cleaned up following the testing to avoid disrupting production operations. The impact of testing failovers that include application recovery can be mitigated with adequate disk array and server sizing.
- When an environment is significantly modified or reconfigured, initiate a test failover to ensure that the recovery plans remain fully operational.
- Because both unplanned failovers and test failovers leave databases in a *crash-consistent* state, some number of uncommitted transactions exist that were not written to the database files. This requires a data consistency check (soft recovery) when the VMs and applications are re-started during a failover. The resulting application recovery intricacies (transaction log replay, database consistency checks, and so forth) can generate a significant amount of additional IOPS, which in turn may affect overall performance. For this reason, HP recommends avoiding peak production hours when executing test failovers.
- Managing the failback process with Site Recovery Manager is a manual process, and the steps vary depending upon the degree of failure at the protected site. For example, if failback follows an unplanned failover, a full data re-mirroring between the two sites may be required.

The combination of understanding all of these major considerations, along with knowing exactly what to do, are keys to the successful deployment of a remote copy infrastructure when using VMware vCenter Site Recovery Manager in conjunction with HP Continuous Access EVA and the EVA VMware Site Recovery Adapter. The test-proven techniques developed here serve as a complete guide that can be used with confidence to ensure success.

For more information

HP solutions and white papers

HP StorageWorks SAN Design Reference Guide
http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00403562/c00403562.pdf

HP StorageWorks Continuous Access EVA
http://h18006.www1.hp.com/products/storage/software/conaccesseva/index.html

HP StorageWorks Continuous Access EVA Implementation Guide
http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01800459/c01800459.pdf

HP StorageWorks Data Replication Designer
http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA1-9942ENW.pdf

HP EVA Performance Analysis using EVAperf
http://h71028.www7.hp.com/ERC/downloads/5983-1674EN.pdf

HP Solutions for VMware
www.hp.com/go/vmware

HP VMware Storage Solution Whitepapers
http://h71028.www7.hp.com/enterprise/us/en/solutions/storage/vmware.html

HP VMware Server Consolidation Solution Whitepapers
http://h71019.www7.hp.com/ActiveAnswers/cache/71088-0-0-0-121.html

VMware references

VMware vCenter Site Recovery Manager 1.0 Evaluator Guide
http://www.vmware.com/pdf/srm_10_eval_guide.pdf

VMware vCenter Site Recovery Manager Product Documentation
http://www.vmware.com/support/pubs/srm_pubs.html

VMware vCenter Site Recovery Manager Compatibility matrix
http://www.vmware.com/pdf/srm_storage_partners.pdf

VMware vCenter Site Recovery Manager Sizing Estimator for SQL Server (to determine Site Recovery Manager backend database requirements)
http://www.vmware.com/files/pdf/Site_Recovery_Manager_1.0U1_Database_Sizing_Calculator_SQL.xls

VMware vCenter Site Recovery Manager Performance and Best Practices for Performance - Architecting Your Recovery Plan to Minimize Recovery Time
http://www.vmware.com/resources/techresources/10057

This document was created using the official VMware icon and diagram library. VMware does not endorse or make any representations about third party information included in this document, nor does the inclusion of any VMware icon or diagram in this document imply such an endorsement.

Technology for better business outcomes

© Copyright 2009 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Copyright © 2009 VMware, Inc. All rights reserved. This document is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at http://www.vmware.com/go/patents.

VMware, VMware vCenter, ESX and VMotion are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions.

Microsoft® and Microsoft Windows® are U.S. registered trademarks of Microsoft Corporation.

4AA2-8848ENW, August 2009